

ONLINE TRACKING AND FIXING OF INVALID GUESS-DBAs
IN SECONDARY INDEXES AND MAPPING TABLES
ON PRIMARY B+TREE STRUCTURES

5

Field of the Invention

The present invention relates to organizing and accessing data in a database using database indexes. In particular, the present invention relates to auxiliary structures such as mapping tables and secondary index structures for indexing tables stored as primary B+trees. More particularly, the present invention relates to methods and structures for efficient maintenance of mapping tables and secondary index structures for a primary B+tree structure as the primary B+tree structure is updated.

10
15
20

Background of the Invention

In a typical relational database system, users store, update, and retrieve information by interacting with user applications. The applications respond to a user's interaction by submitting commands to a database application, or server, responsible for maintaining the database. The database server responds to commands by performing the specified actions on the database. To be correctly processed, the commands must comply with the database language that the database server supports. One popular database language is commonly known as Structured Query Language (SQL).

Various access methods may be utilized to retrieve data from a database. The access methods used to retrieve data may significantly affect the speed of the retrieval and the amount of resources consumed during the retrieval process. Many information retrieval applications make use of indices when performing content-based searches on the database data. Examples of 5 database indices include R-trees, quadtrees, and B-trees.

Database indices provide organization and reference to the data in a database to permit a user to find particular items of data in the database or determine relationships among the data in the database. Database indices can also permit relationships between the data in a database and data not included in the database to be determined. For example, an index can make it possible to determine location within a certain distance of a location defined in a database comprised of geographical location information.

Summary of the Invention

The present invention addresses problems associated with block addresses stored in secondary indexes and/or mapping tables for a primary B+tree structure becoming stale due to row movement in the primary B+tree structure caused by leaf block splits. No known solution exists for addressing the problems that the present invention resolves.

20

The present invention provides a method for maintaining a system for database management. The method includes recording an address of a newly created block resulting from splitting of a leaf block of a primary B+tree and maintaining the address in a list as part of

primary B+tree metadata.

Additionally, the present invention concerns a system for organizing a database index.

The system includes a list as part of primary B+tree metadata. The list maintains addresses of
5 newly created leaf blocks during split operation of the primary B+tree.

Also, the present invention relates to a computer program product for performing a process for maintaining a database management system. The computer program product includes a computer readable medium and computer program instructions recorded on the computer readable medium and executable by a processor. The computer program instructions perform steps including recording a new address for a newly created block during splitting of a leaf block of a database and maintaining the new address in a list as part of primary B+tree metadata.

Furthermore, the present invention provides a system for performing a process for maintaining a database management system. The system includes a processor that is operable to execute computer program instructions. The system also includes a memory operable to store computer program instructions executable by the processor. The computer program instructions perform steps including recording a new address for a newly created block during splitting of a leaf block of a database and maintaining the new address in a list as part of primary B+tree metadata.

Still other objects and advantages of the present invention will become readily apparent

by those skilled in the art from a review of the following detailed description. The detailed description below shows and describes preferred embodiments of the present invention, simply by way of illustration of the best mode contemplated of carrying out the present invention. As will be realized, the invention is capable of other and different embodiments and its several details are capable of modifications in various obvious respects, without departing from the invention. Accordingly, the drawings and description are illustrative in nature and not restrictive.

Brief Description of the Drawings

Objects and advantages of the present invention will be more clearly understood when considered in conjunction with the accompanying drawings, in which:

Fig. 1 represents a block diagram of an embodiment of a database management system according to the present invention.

Detailed Description of the Invention

For the primary B+-tree structures, an additional mapping table can be created as
09/867,678
described in a U.S. patent application filed on even date herewith for "Mapping Logical Row
Identifiers For Primary B+-Tree-Like Structures To Physical Row Identifiers", to Chong et al.,
and having attorney docket number 19111.0038, to supporting bitmap indexes, which are
number 09/865,599 now abandoned
described in a U.S. patent application filed on even date herewith for "Supporting Bitmap

Indexes on Primary B+tree Structures", to Chong et al., and having attorney docket number 19111.0040, the entire contents of the disclosures of both of which are hereby incorporated by reference. Also, bitmap indexes are described in greater detail in U.S. patent 5,363,098, for

"Byte Aligned Data Compression", issued November 8, 1994, to Antoshenkov, the entire

5 contents of the disclosure of which are hereby incorporated by reference. Additionally or alternatively, secondary indexes can be created with logical row identifiers as described in U.S. ^{now U.S. Patent 6,546,394} patent application serial number 09/473,073[^], to Chong et al., filed December 28, 1999, for "Database System Having Logical Row Identifiers", the entire contents of the disclosure of which is hereby incorporated by reference. Both mapping tables, which may be utilized for supporting bitmap indexes, and secondary indexes store physical data block address to speed up query performance.

10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635
8640
8645
8650
8655
8660
8665
8670
8675
8680
8685
8690
8695
8700
8705
8710
8715
8720
8725
8730
8735
8740
8745
8750
8755
8760
8765
8770
8775
8780
8785
8790
8795
8800
8805
8810
8815
8820
8825
8830
8835
8840
8845
8850
8855
8860
8865
8870
8875
8880
8885
8890
8895
8900
8905
8910
8915
8920
8925
8930
8935
8940
8945
8950
8955
8960
8965
8970
8975
8980
8985
8990
8995
9000
9005
9010
9015
9020
9025
9030
9035
9040
9045
9050
9055
9060
9065
9070
9075
9080
9085
9090
9095
9100
9105
9110
9115
9120
9125
9130
9135
9140
9145
9150
9155
9160
9165
9170
9175
9180
9185
9190
9195
9200
9205
9210
9215
9220
9225
9230
9235
9240
9245
9250
9255
9260
9265
9270
9275
9280
9285
9290
9295
9300
9305
9310
9315
9320
9325
9330
9335
9340
9345
9350
9355
9360
9365
9370
937

the rows move may be recorded. The DBA of that block may be maintained in a list as part of primary B+tree metadata. Also, a count of DBAs in the list may be maintained.

In order to determine whether it is necessary to carry out the present invention, a measure 5 of invalid guess-DBAs may be maintained. This ratio may be utilized to adjust the guess-DBA quality of mapping table and/or secondary indexes. According to one embodiment, the measure includes the ratio <count of DBAs>/<total # of leaf blocks> as a measure of invalid guess-DBAs in mapping table and/or secondary indexes. According to the present invention, the list of DBAs is maintained only when the ratio is less than a predetermined threshold, such as, for example, about 10%.

At the point the selected measure described above falls below the predetermined threshold, or any measure indicates that the quality of the mapping table, secondary index or other structure has fallen below a certain level, the present invention can include initiating operation, either explicitly by user or implicitly by the system, for revising the mapping table, secondary index and/or other structure. Typically, the revision includes a selective fix-up of corresponding mapping table and secondary index entries.

The following describes one embodiment for carrying out the revisions in a mapping 20 table. For all rows in the list of blocks, to fix mapping table entries, the values for the following elements may be retrieved from the primary B+tree: corresponding mapping table row identifiers and the DBA of the current block in the list. Next, these values may be sorted in the order of mapping table row identifiers. Then, mapping table rows corresponding to the mapping

table row identifiers may be retrieved and its guess-DBA component is updated, if it differs from the current DBA.

INS. *AI* In the context of a secondary index, entries in the index may be updated as described below. For all rows in the list of blocks, to fix secondary index entries, the values for the following elements may be retrieved from the primary B+tree: a secondary index key, a primary key, and a DBA of the current block in the list. Then, the values may be sorted in order of the (secondary index key, primary key) pairs. Next, an index row corresponding to the (secondary index key, primary key) pair may be retrieved and its guess-DBA component is updated if it differs from the current DBA.

The updating of a mapping table and/or a secondary index structure may be carried out on-line. Alternatively, the updating may be carried out off-line. Regardless of when the updates are carried out, they may be carried out in a piece-wise manner, particularly when only the guess-DBA is being updated. Similarly, regardless of when the updating occurs, the updates may be committed in batches.

The above discussion of updating relates to situations where the defined ratio falls below a threshold value. On the other hand if the measure is above a predetermined threshold, it may be desired to carry out other operations on guess-DBAs.

The following describes an embodiment of a process for resorting of guess-DBAs to a per object based fixing. For each mapping table or secondary index a full scan of the object may be

performed. For each row of the mapping table or secondary index, the correct guess-DBA may be determined by traversing the primary B+tree structure only up to the penultimate level. Then, each row of the mapping table or secondary index may be updated with the correct guess-DBA. Subsequently, the correct guess-DBAs may be committed to the mapping table or secondary index in small batches.

To permit the above-described methods to be carried out, the present invention also includes a system for organizing databases. The system includes a list of addresses of block newly created during split operation on the primary B+tree. Typically, the auxiliary structures that benefit from this invention are a mapping table and/or a secondary structure. However, the present invention could be utilized with other structures. The system according to the present invention can also include a count of database block addresses in the list. Additionally, the system can also include a ratio of database block addresses to total number of leaf blocks as a measure of invalid guess-DBAs.

An exemplary block diagram of a database management system 100 according to the present invention is shown in Fig. 1. A database management system typically includes a programmed general-purpose computer system, such as a personal computer, workstation, server system, and minicomputer or mainframe computer. The embodiment of the database management system 100 shown in Fig. 1 includes processor (CPU) 102, input/output circuitry 104, network adapter 106, and memory 108. CPU 102 executes program instructions in order to carry out the functions of the present invention. Typically, CPU 102 is a microprocessor, such as an INTEL PENTIUM® processor, but may also be a minicomputer or mainframe computer

processor.

Input/output circuitry 104 provides the capability to input data to, or output data from, computer system 100. For example, input/output circuitry may include input devices, such as 5 keyboards, mice, touchpads, trackballs, scanners, etc., output devices, such as video adapters, monitors, printers, etc., and input/output devices, such as, modems, etc. Network adapter 106 interfaces database management system 100 with network 110. Network 110 may be any standard local area network (LAN) or wide area network (WAN), such as Ethernet, Token Ring, the Internet, or a private or proprietary LAN/WAN.

10
11
12
13
14
15
16
17
18
19

20

Memory 108 stores program instructions that are executed by, and data that are used and processed by, CPU 102 to perform the functions of the present invention. Memory 108 may include electronic memory devices, such as random-access memory (RAM), read-only memory (ROM), programmable read-only memory (PROM), electrically erasable programmable read-only memory (EEPROM), flash memory, etc., and electro-mechanical memory, such as magnetic disk drives, tape drives, optical disk drives, etc., which may use an integrated drive electronics (IDE) interface, or a variation or enhancement thereof, such as enhanced IDE (EIDE) or ultra direct memory access (UDMA), or a small computer system interface (SCSI) based interface, or a variation or enhancement thereof, such as fast-SCSI, wide-SCSI, fast and wide-SCSI, etc, or a fiber channel-arbitrated loop (FC-AL) interface.

INS *A2* Memory 108 includes a plurality of blocks of data, such as new address block 112, list block 114, and ratio block 116, and a plurality of blocks of program instructions, such as

processing routines 118 and operating system 120. New address block 112 stores a plurality of new addresses for rows split from a leaf block that have been received by the database management system 100 as a database index is modified. List block 114 stores a list of the new addresses as metadata for the database index. Ratio block 116 stores the ratio of database addresses to total number of leaf blocks that may be used to evaluate invalid guess-DBAs.

5 Processing routines 118 are software routines that implement the processing performed by the present invention. Operating system 120 provides overall system functionality.

10 It is important to note that while the present invention has been described in the context of a fully functioning data processing system, those of ordinary skill in the art will appreciate that the processes of the present invention are capable of being distributed in the form of a computer readable medium of instructions and a variety of forms and that the present invention applies equally regardless of the particular type of signal bearing media actually used to carry out the distribution. Examples of computer readable media include recordable-type media such as floppy disc, a hard disk drive, RAM, and CD-ROM's, as well as transmission-type media, such as digital and analog communications links.

15

20 The foregoing description of the invention illustrates and describes the present invention. Additionally, the disclosure shows and describes only the preferred embodiments of the invention, but as aforementioned, it is to be understood that the invention is capable of use in various other combinations, modifications, and environments and is capable of changes or modifications within the scope of the inventive concept as expressed herein, commensurate with the above teachings, and/or the skill or knowledge of the relevant art. The embodiments

described hereinabove are further intended to explain best modes known of practicing the invention and to enable others skilled in the art to utilize the invention in such, or other, embodiments and with the various modifications required by the particular applications or uses of the invention. Accordingly, the description is not intended to limit the invention to the form disclosed herein. Also, it is intended that the appended claims be construed to include alternative embodiments.

CONFIDENTIAL INFORMATION